Searching for Hotspots

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAR 2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Searching for Hotspots	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Neptune and Company, Inc,1435 Garrison Street, Suite 110,Denver,CO,80215**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the 9th Annual DoD Environmental Monitoring and Data Quality (EDMQ) Workshop Held 26-29 March 2012 in La Jolla, CA. U.S. Government or Federal Rights License

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**16**	

Isn't there an easy formula?

If we are estimating a mean:

$$n = \frac{(z_{1-\alpha/2})^2 s^2}{d^2} + \tfrac{1}{2} z_{1-\alpha/2}^2$$

If we are estimating a mean to make a decision:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 s^2}{d^2} + \tfrac{1}{2} z_{1-\alpha}^2$$

and using either requires a random sample…

…but these are for means (averages)

- The formulae for estimating a mean will not work if the objective is the detection of the presence of a potential "hot-spot“.

- A different approach is needed, one that uses a grid to locate samples and a threshold value such that above this value implies a hotspot.

- The problem becomes:
  - How big is the potential hotspot
  - How large a mesh-spacing is needed
  - What is the chance of missing a hotspot

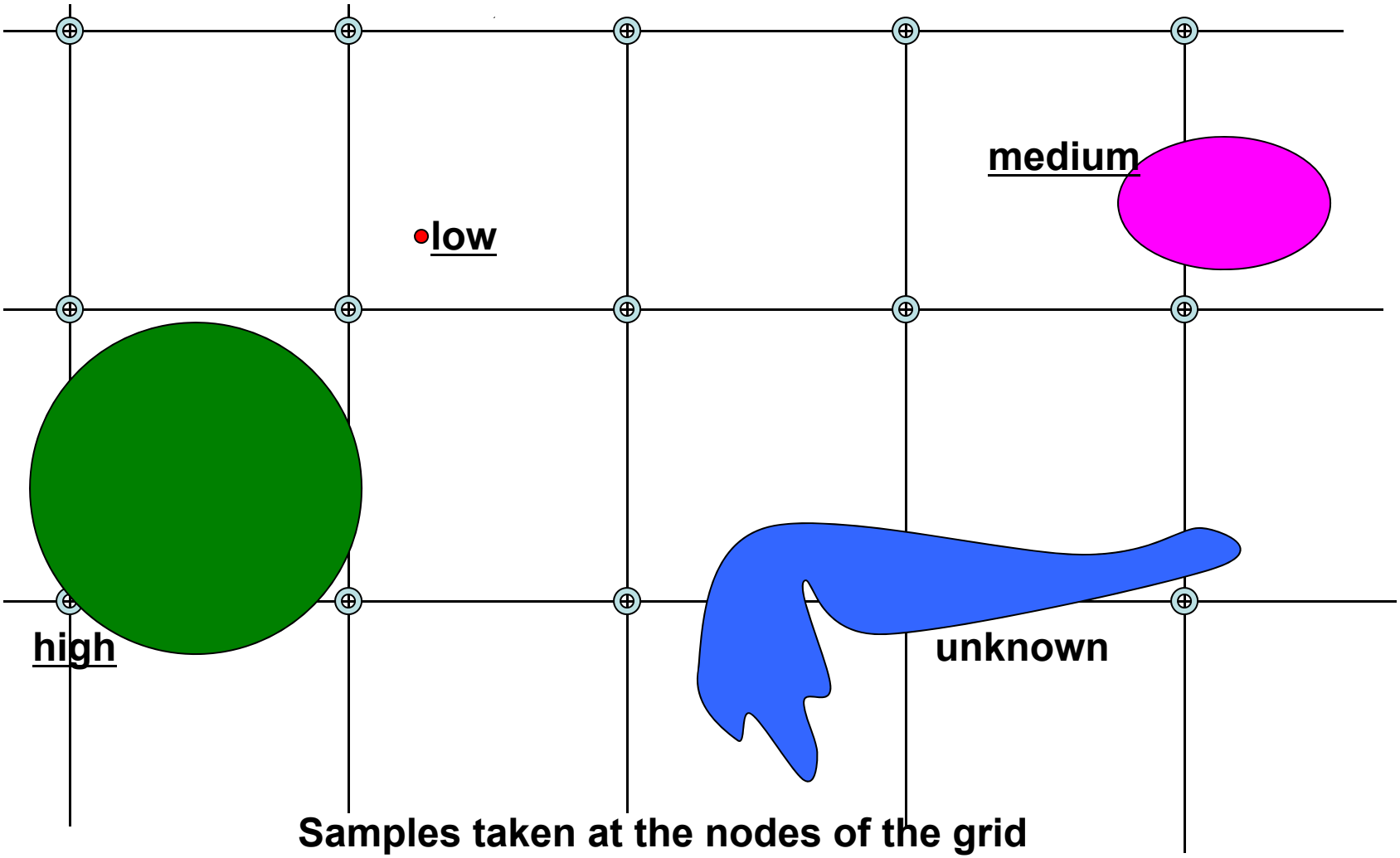
Specify one of the three

- How big is the potential hotspot?
- If hotspot is bigger than the mesh, then 100% chance!
- If very small then very low chance of catching it

- How large a mesh-spacing is needed?
- If large spacing used, then high risk of missing hotspot
- If small spacing, then will be collecting many samples

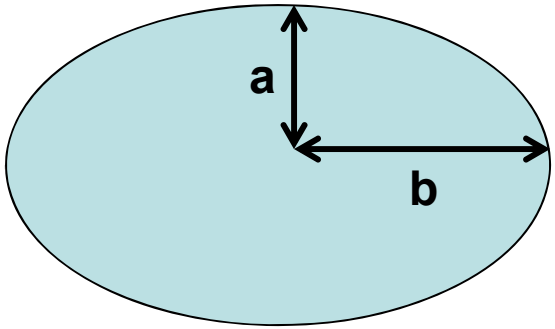
- What about the chance of missing a hotspot?
- Depends on how certain do you want to be
- If its important to find then will need many samples

Hot-Spot size, grid spacing, and probability of missing



How hot-spot sampling works

- You specify the size and decide if it is circular (radius "a") or elliptical (short arm "a", longer arm "b").

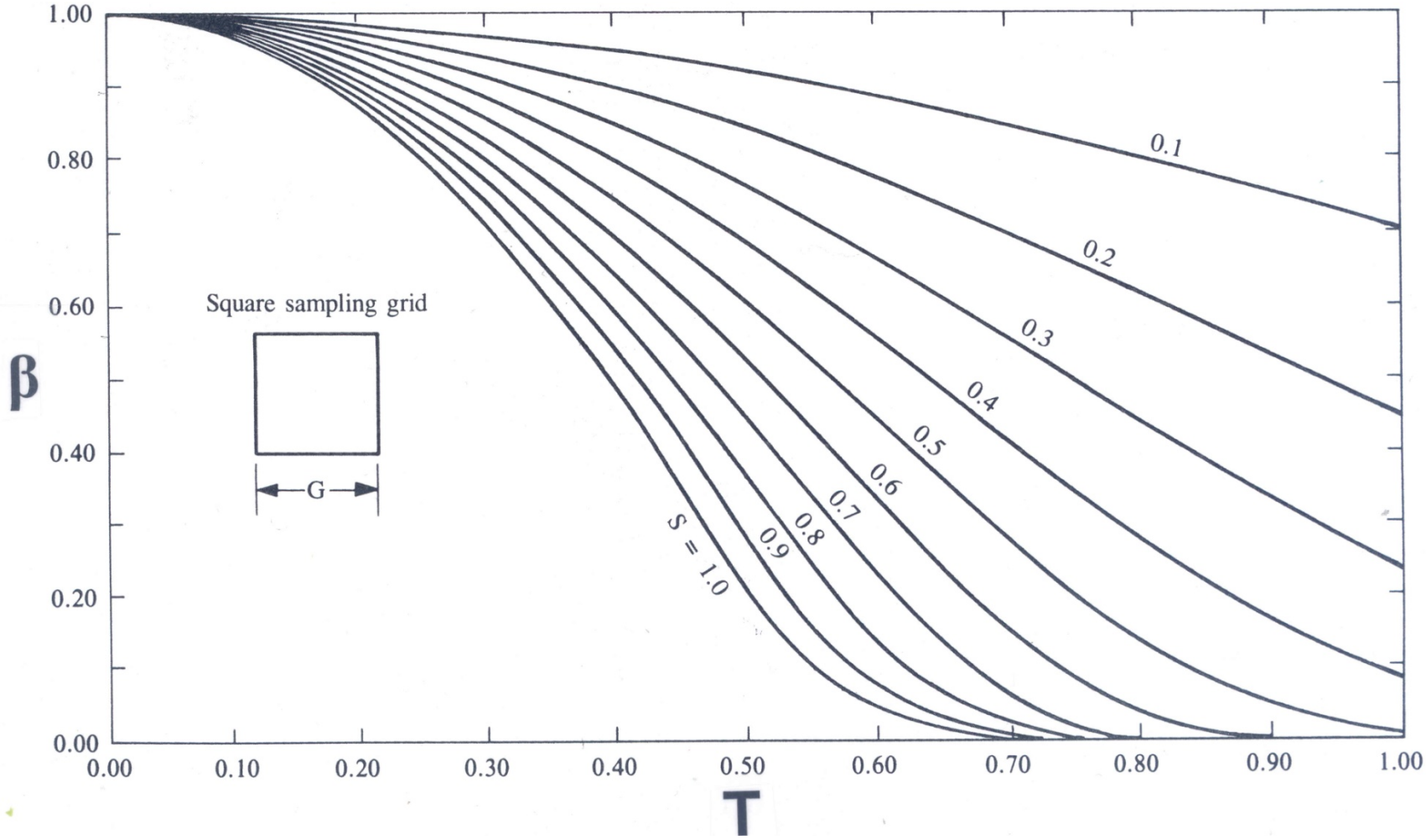


Calculate S = a/b

- You decide how much risk you are willing to take on missing hitting the hot-spot (the risk, probability "β").

- You use "β" and "S" with the following nomogram to find T, then the Grid mesh size, G = b/T.

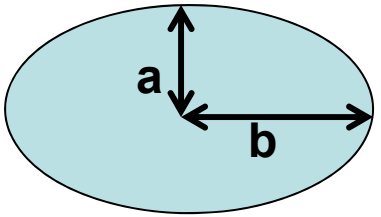
Hot-spot nomogram



Example: Hot-spot grid size calculation

4 acre abandoned property thought to have dispensed pesticides from a single location over 10 years ago. No information on probable location of filling area. 4 acres = roughly 4 football fields.

- **You specify the size and shape:**
  Elliptical, a=20ft, b=40ft



- **You decide risk of missing hitting hot-spot:**
  No more than a 10% chance (β = 0.10)

- **Calculate S = a/b: S = 20/40 = 0.50**

- **Use "β" and "S" with the nomogram to find T: T = 0.82**

- **Grid mesh size G = b/T: G = 40/0.82 = 48.8 ft**

Hot spot sampling (cont)

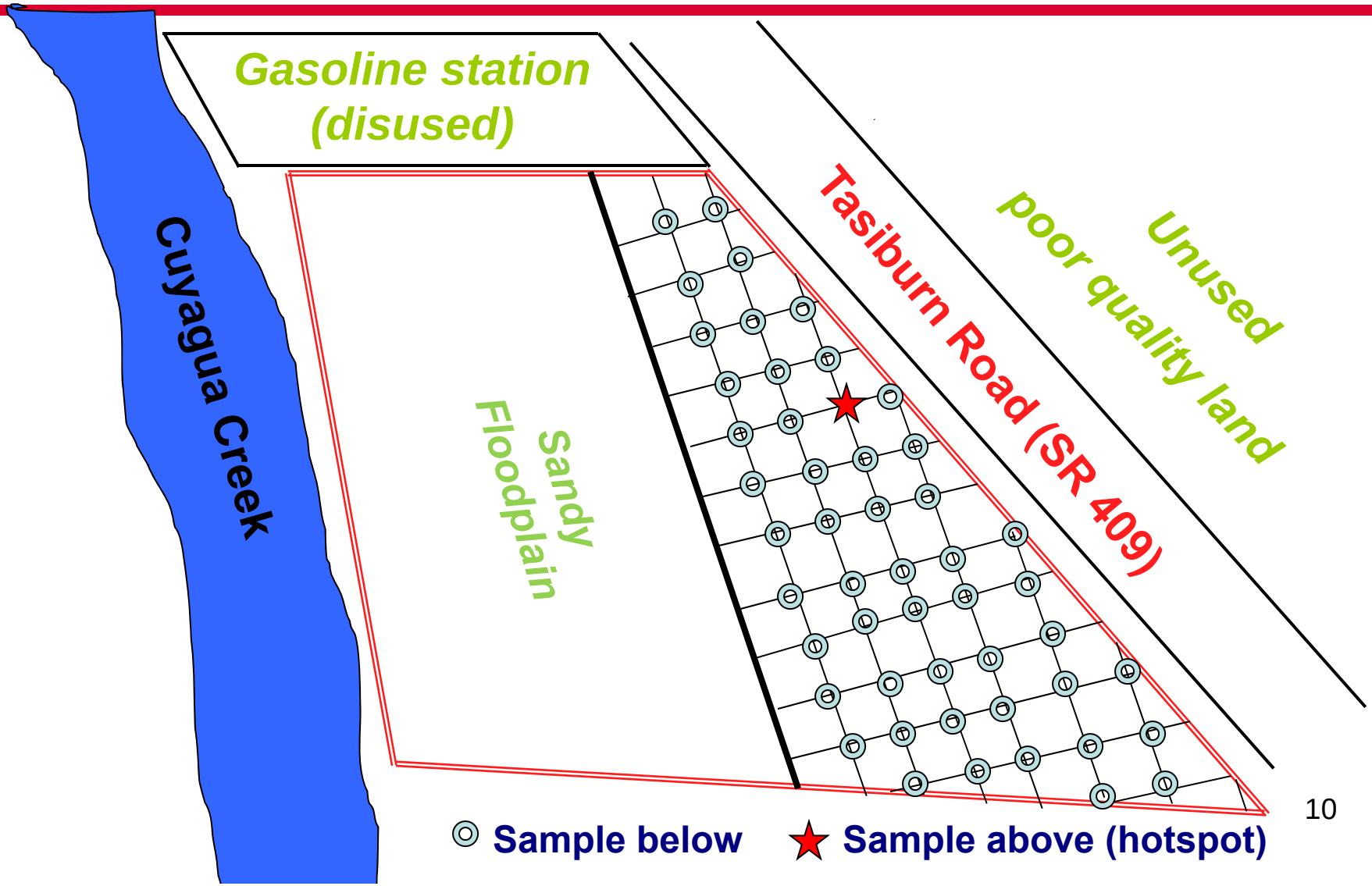
- Now asking a field crew to create a grid with a mesh size 48.8ft (48ft 9½in) is very difficult.

- Working backwards from the nomogram:
  - If 50ft mesh selected then $\beta = 0.17$ (17%)
  - If 45ft mesh selected then $\beta = 0.08$ (8%)

- Original requirement was β not to exceed 10% so a choice has to be made:

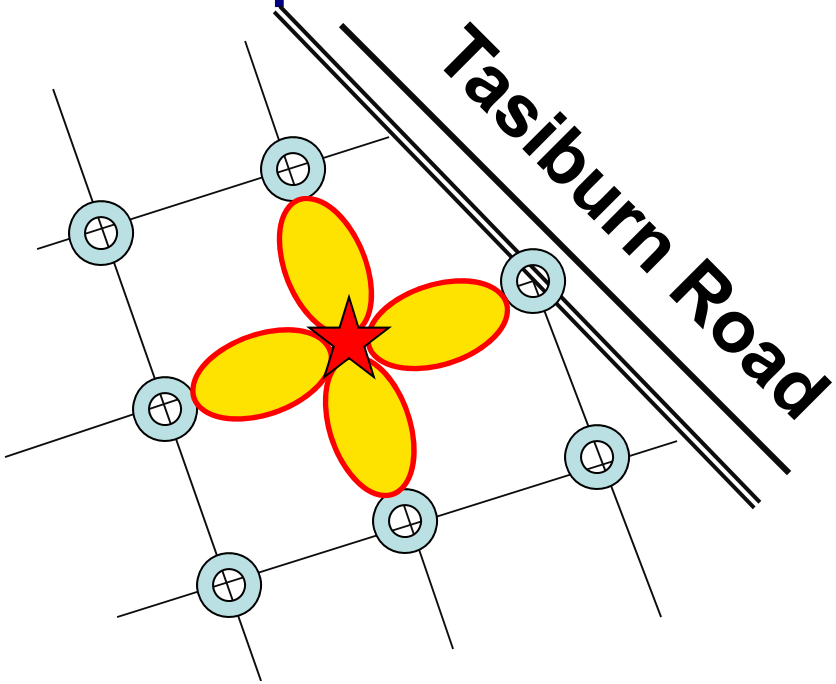
Select 50ft mesh with higher chance of missing hotspot or select 45ft mesh but take more samples?

Samples results compared to a threshold



Conclusions from the example

- Using a 50ft grid mesh we covered the site with 50 samples taken at the nodules of the grid
- 1 sample was above the threshold thus implying the potential existence of a hotspot

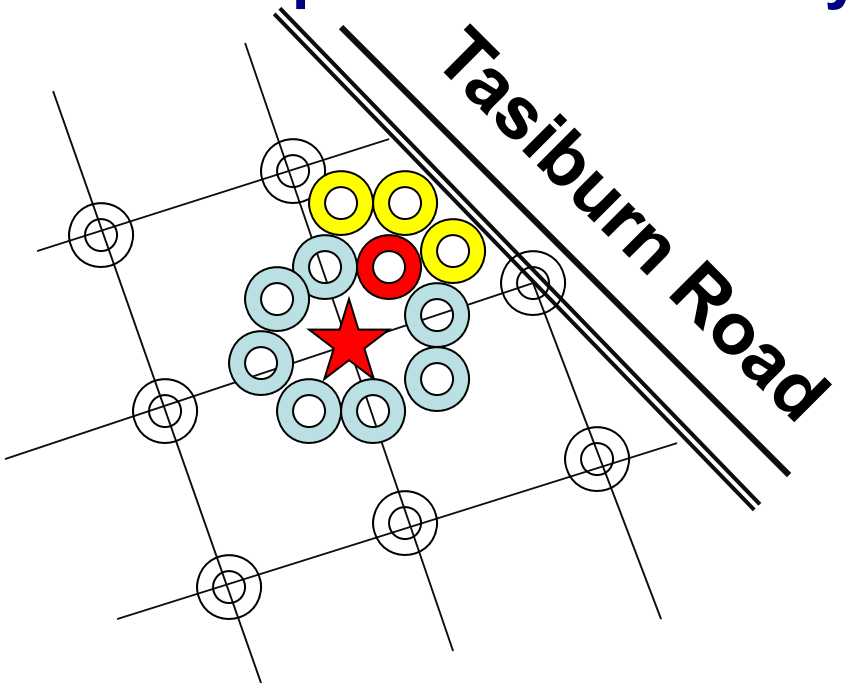


Hotspot

Potential extent of hotspot

Further sampling to determine the extent of the hotspot

- Using the position of the as the central location, further samples are taken around this position
- This is repeated until the boundaries of the hotspot are sufficiently delineated



Hotspot

Sample above threshold

Sample below threshold

Potential further samples

Hot Spot Sampling: Pro

- Very effective in finding hot spots of a specified size and shape
- Able to control the probability of missing a hot-spot of a given size
- Indirectly delineates the size and shape of the hot-spot
- Results can be used to guide further investigation

Hot Spot Sampling: Con

- Need to know what size and shape the potential hot-spot should be
- Need to decide the chance of missing a hot-spot of a certain size
- Could be expensive as you need to take samples from all over the grid

What can we infer (qualitatively)?

- Samples taken across the entire grid revealed a hot-spot, indicated by a red star 
- After identification of the hot spot, 8 extra samples were taken around the hot-spot but off the grid (recall initial samples were from the nodes of the grid mesh)
- 7 of the extra samples were considered "not hot"  , but 1 sample could be considered "hot"  , further samples could delineate the extent of the hotspot
- The hotspot is in the form of a plume stretching from Tasiburn Road roughly 60 ft into the field

Can we increase the quantitative information using extra samples?

- **No.** The new data is not random (like the original data) but deliberately selected as there was a hot-spot guiding them. You can't just add them together and expect the inferences (probabilities) to come out right.

- **Yes.** However, it is not straight-forward. This kind of sampling is known as Adaptive Cluster Sampling and you have to use special formulae to adjust for the fact you're using non-random samples.